DOCUMENT RESUME

ED 309 760 IR 013 926

AUTHOR Wise, Steven L.; Kutish, Gerald W.
TITLE Data Desk Professional: Statistical Analysis for the Macintosh.
PUB DATE Mar 89
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 25-30, 1989).
PUB TYPE Book/Product Reviews (072) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Software; Computer Software Reviews; *Microcomputers; *Statistical Analysis
IDENTIFIERS Apple Macintosh; Data Desk Professional; *Exploratory Data Analysis

ABSTRACT

This review of Data Desk Professional, a statistical software package for Macintosh microcomputers, includes information on: (1) cost and the amount and allocation of memory; (2) usability (documentation quality, ease of use); (3) running programs; (4) program output (quality of graphics); (5) accuracy; and (6) user services. In conclusion, it is noted that the program is oriented toward an exploratory data analysis (EDA) approach, which encourages an open-ended analysis of data, and that the user is able to examine a data file in a very flexible and thorough fashion. In terms of traditional statistical analysis, however, Data Desk has a number of clear deficiencies, in that some of the programs are inflexible and incomplete. (GL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED309760

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Data Desk Professional: Statistical Analysis for the Macintosh

Steven L. Wise

Gerald W. Kutish

University of Nebraska-Lincoln

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, March, 1989

IR013926

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Steven L. Wise

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Data Desk Professional: Statistical Analysis for the Macintosh

As microcomputers have grown in sophistication, so too have available statistical software packages. This paper provides a review of Data Desk Professional, an integrated data analysis system designed for Macintosh microcomputers. The review of Data Desk will be structured around the microcomputer statistical software review model developed by Ansorge, Wise, and Plake (1987).

General Information

Data Desk, a product of the Odesta Corporation in Northbrook, Illinois, is currently available at a price of $247. It requires at least 512K of RAM, and it allocates memory dynamically, trading space among data, program, and results as needed. Each Data Desk data file has a generous limit of 500 variables. Moreover, variables can have up to 32,000 cases, although users working on a Macintosh Plus or SE are urged to limit their data files to a few thousand cases. Hence, for most applications, space limitations should not pose problems to users. A smaller student version of Data Desk is available; users of this package are limited to data files containing a maximum of 15 variables and 1000 cases. A copy shop near the University of Nebraska campus currently sells this student package for $38.

A primary feature of Data Desk is the ease with which a user can engage in Exploratory Data Analysis (EDA) (Tukey, 1977). The program documentation encourages users to explore their data in an open-ended fashion, without a priori decisions regarding the data

analyses to be made. A clear major goal of the authors of Data Desk was to develop a flexible, sophisticated EDA system. The Data Desk Handbook states that

> Data Desk differs from traditional statistics packages because it is designed for the entire process of data analysis rather than for the computing of specific statistics. Data analysis is the process of discovering, describing, and confirming structure or patterns in data. The process itself is often a way to learn about data rather than being just a means of obtaining final, clean results. (pp. 2/1)

Data Desk, however, can also readily be used to perform more traditional, hypothesis-driven data analysis.

Software Usability

A major aspect of statistical software quality is the extent to which the user's data analysis needs are met. While it is important to evaluate statistical software in terms of the programs it can run and the impressiveness of the displayed output, the "usability" of a software program is additionally a function of many diverse dimensions such as clarity of documentation, accuracy of algorithms, and availability of user services. A variety of these dimensions of usability will be discussed below.

Documentation

Three basic manuals accompany the Data Desk program disks. The Quickstart Guide provides an introduction to the system and leads the user through an EDA-type data analysis session using one of the sample data files included with the programs. This introductory manual is well written and should be quite useful to new users. The

Handbook manual provides a comprehensive guide to understanding and operating the Data Desk system. The material in this manual is presented in a clear, consistent manner. In particular, extensive manual space is devoted to entering and editing data, as well as importing and exporting data files. The Statistics Guide provides detailed descriptions of the statistical procedures. Along with the description of each procedure, most of the formulas being used are displayed and discussed. Thus, a user has a good idea of how a procedure is operating and can decide if the algorithm is appropriate for his/her needs. A notable exception is the absence of a detailed explanation of how an ANOVA design with unequal sample sizes is analyzed.

Ease of Use

Learning to use all of the features of Data Desk takes some time. The new user is faced with the task of learning much terminology and many concepts that are specific to this software system. Numerous terms such as "bundles", "HyperViews", and "editing sequences" need to be mastered in order to effectively use the system. Data Desk allows the user much power and flexibility in displaying and understanding his/her data; consequently, the system is quite complex and is a bit more difficult to learn than many other statistical software systems. However, given the highly complex nature of Data Desk, the developers have done a commendable job of making the system as easy to use as possible. A large set of sample data files are included to help users learning to use the programs. When using the system, the user's choices are exclusively menu driven, and the user can jump to a help

file at any time. In addition, the error diagnostics are clear and informative.

Running Programs

The data entry procedures are both flexible and easy to use. Data can be entered for a single variable or for several variables at a time (e.g., all of the data for a case). Editing, modifying, or transforming variables is very easy to accomplish. A wide variety of data files can be imported, including spreadsheet, database, and word processor files.

Depending on the user's strategy for analyzing his/her data, the statistics capabilities of Data Desk range from superb to disappointing. Users who approach data analysis from an EDA perspective will be very excited about Data Desk's capabilities to sift through their data, allowing them to study the emerging relationships and to focus the analysis on subsets of the data. This aspect of Data Desk is outstanding.

On the other hand, users who wish to conduct more traditional hypothesis-driven analyses are apt to be disappointed. Many of the programs are deficient in their output. For example, when displaying a scatterplot, running a simple linear regression does not produce a graph of the regression line. Moreover, the overall model tests and the tests of the regression weights do not provide accompanying probability levels. Similarly, the program that computes a correlation matrix provides no test of the significance of the correlations.

Most of the statistical programs could be improved, some substantially. The analysis of variance program is only available for standard, fixed-effects models. The ANOVA output lacks information from which a user can assess effect size, such as R^2, coefficient of

variation, or root mean square. Tests of model assumptions are not available, nor are follow-up multiple comparison (e.g., Tukey) tests. For factorial designs, significant interactions can not be followed up using tests of simple effects, a common method for analyzing interactions. In short, the ANOVA program is inflexible and incomplete.

Two other programs deserve special mention. First, the cluster analysis program is rudimentary and relatively uninformative. Using Euclidean distance as the only available proximity measure is very restrictive. The graphic display of the output is difficult to assimilate, since the linkage scale is not displayed, and tests of goodness of fit are not provided. Second, the principal components program provides little beyond the eigenvalues, eigenvectors, and the unrotated factor matrix. It is difficult for the user to go beyond this information. One can neither select only a portion of the components nor rotate the factor matrix.

We have described only some of the deficiencies that we have identified in the programs. Someone who is considering purchasing Data Desk would be well advised to check that the programs are adequate for his/her statistical needs. Specifically, users who wish to conduct relatively sophisticated analyses may be unhappy with Data Desk's capabilities.

Program Output

In general, the output provided by Data Desk is clearly and logically displayed. The graphics are satisfactory, although they are relatively sterile in appearance. As mentioned earlier, however, there is numerous information missing from the output.

Accuracy

This is a crucial aspect of a statistical software review. The accuracy of many of the programs was checked by comparing the Data Desk results with those obtained using both SAS and SPSSX. In every comparison, the results were virtually identical; agreement was found to at least 8 significant digits of computation and from 4 to 8 significant digits in the output. It appears that users of Data Desk can be confident that its programs will yield results that are as accurate as those obtained using a large statistical package on a mainframe computer.

User Services

It is important that users of statistical software have resources beyond the programs and documentation. The typical user will inevitably encounter problems that cannot be solved by reference to the program documentation. A toll-free telephone number for user assistance is provided in the "About Data Desk" selection on the Apple menu. We have not used this number and hence cannot evaluate the usefulness of this number in obtaining assistance. We are unaware of any additional user services such as training sessions or a user newsletter.

Recommendations and Conclusions

Data Desk is a very advanced statistical software package that takes advantage of many features of the Macintosh. Oriented toward an EDA approach to data analysis, it allows a user to examine a data file in a very flexible and thorough fashion. In terms of traditional statistical analysis, on the other hand, Data Desk has a number a clear

deficiencies. It might be argued that EDA should be an initial step in most data analyses. The fact remains, however, that most data analysts do not engage in EDA. Hence, for these people, Data Desk may not be the most useful statistical analysis package available for the Macintosh.

The only real areas of weakness in Data Desk lie in the statistics programs. If these programs were to be substantially improved, the quality of the system would be greatly enhanced. Until such improvements occur, Data Desk will continue to be of limited use to data analysts who do not approach data analysis from an EDA perspective.

References

Ansorge, C. J., Wise, S. L., & Plake, B. S. (1987). Evaluating the quality of microcomputer statistical software. Computers in Human Behavior, 3(1), 73-82.

Tukey, J. W. (1977). Exploratory data analysis. Reading, Mass.: Addison-Wesley.